A Fuzzy-Based Rough Sets Classifier for Forecasting Cellular Mobile Call Abandon Rate

A.C. Onwuekwu[1], S.M. Sani[2], M.B. Mu'azu[3]

[1,2,3](Department of Electrical and Computer Engineering, Ahmadu Bello University, Zaria, Nigeria)

Abstract: *Abandon Rate is an effective evaluation indicator for telecommunication operators to measure network performance. It is the percentage of incoming calls made to a call or service center that are dropped by the callers before they are being attended to by a call center representative (agent). The research focuses mainly on improving abandon rate forecasting problems and offers an intelligent hybrid model in which decision rules will be generated as programs that will be able to reason and use a knowledge base to classify and forecast abandon rate. This proposed hybrid model is constituted by five artificial intelligence (AI) tools: experiential knowledge (EK), Discretization method - Minimize Entropy Principle Approach (MEPA), fuzzy set theory (FST), rule filter (RF), and rough set theory (RST). The model will be validated using obtained data from Airtel call center collected over a three months period. The proposed model is expected to improve forecasting performance in terms of higher accuracy, using fewer attribute sets in rule generation, and generates fewer numbers of decision rules.*
Keywords: *Abandon rate, Classifier, Experiential knowledge, LEM2 algorithm, MEPA.*

I. Introduction

The increasing demand in acquiring information and rapid growth in technology brought about the increase in use of mobile handsets in almost all parts of the world. The telecommunication industry is changing very rapidly due to advance in technology. According to various research works, the number of mobile phone users is increasing by the day and this increase result to serious market competition among the service providers [1]. Call centers are places or offices set up by companies to receive calls from customers in order to assist them with queries or give information about a product or service. Companies are able to communicate with their customers through the call centers. One aim of telecommunication operators is to offer good services to the customers, not only to make profit, but also to retain customers' loyalty. This can be achieved by avoid losing large number of total calls offered into the centers.

Large numbers of call centre representatives are scheduled hourly to attend to large number of phone callers in very large call centers. Interactive Voice Response (IVR) units, which are specialized computers that allow customers to perform some self-service transactions, are used in many inbound call centers. If the customer still wants to speak with an agent, he is connected to any available agent; or he joins a queue of waiting callers. The customer is eventually served by an available agent or becomes impatient or tired of waiting on the queue and hangs up before an agent is made available. The call center can be modeled as a queuing system with this description [2]. The longer the waiting time those customers have to wait on queue before they are being connected to speak to an agent, result to high abandon rates. Forecasting is performed by service providers to assist them in planning and evaluating their networks, and in budgeting.

Earlier machine learning and forecasting methods like statistical methods have limitations which include linear separability of sets, multivariate normal or Gaussian distribution and the predictive variables independence. With these limitations, telecommunication operators require more efficient approaches to forecast call abandon rate accurately and to improve call center performance. There is importance to understand the call traffic volume at any given period and also to understand customer abandonment or patience behavior. Various research works have been carried out on artificial intelligence techniques for classification. These techniques include; rough set theory (RST), fuzzy set theory (FST), decision trees (DT) and artificial neural networks (ANN). The rapid change in information technology resulted in these research trend [3]. An intelligent hybrid model is proposed which is based on rough set theory to generate decision rules. The hybrid model is constituted of five AI components and they are: experiential knowledge, discretization method (MEPA), fuzzy set theory, rough set theory and rule filter, for forecasting abandon rates using inbound call data from Airtel call center, Abuja. The proposed hybrid model is aimed at improving forecasting accuracy.

II. Theoritical Background

An overview of concepts and AI components required for the proposed hybrid model will be discussed in this section. These include of rough set theory, LEM2 algorithm, rule filter, fuzzy logic and Minimize Entropy Principle Approach (MEPA).

2.1. Abandon Rate

Abandon rate is the percentage of incoming calls made to a call center that are abandoned by the caller before being attended to by a call center representative. It is one of the effective evaluation indicator used by telecommunication operators to measure system congestion. Abandon rate is obtained using the formula:

$$\left(\frac{\text{Number of Calls Abandoned}}{\text{Total Number of Incoming Calls}}\right) \times 100 \tag{1}$$

2.2. Rough Set Theory (RST)

Rough set theory was first proposed in 1982 by Pawlak to deal with problems with uncertainty and vagueness in information systems analysis. It uses specific criteria of the equivalence class concept to partition data sets as belonging to a upper or lower approximation of the original set. In RST, certain set boundaries are refused, and every set will be defined by the lower and upper approximation [4]. Some basic key definitions of RST are given. An analyzed dataset can be represented as information systems.

Let $S = (U, A)$, be an information system, $X \subseteq U$ and $B \subseteq A$ be subsets of the information system. Where U is a set of objects and A is a set of attributes. Both sets are non-empty. To every $X \subseteq U$, two sets called the B-lower and the B-upper approximation of X denoted as $\underline{B}(x)$ and $\overline{B}(x)$, respectively, are defined as follows[5]:

$$\underline{B}(x) = \{x \in U : B(x) \subseteq X\} \qquad \text{(lower)} \tag{2}$$

$$\overline{B}(x) = \{x \in U : B(x) \cap X \neq \varnothing\} \qquad \text{(upper)} \tag{3}$$

The B-lower approximation of a set is the union of all equivalence classes that are included in the set, it includes objects that can be positively classified as belonging to set *X*. Whereas the B-upper approximation of a set is the union of all equivalence classes having non-empty intersections with the set, it includes minimal set of objects that are possibly members of set *X*. The boundary region of set *X* is given by [5]:

$$BN_B(X) = \overline{B}(x) - \underline{B}(x) \tag{4}$$

The boundary region contains inconsistent objects that can neither be classified as being members of or not being members of set *X*. If $BN_B(x) = \varnothing$, (i.e. is an empty set) it indicates that *X* is exact with respect to B; and if $BN_B(x) \neq \varnothing$, (i.e. is not an empty set), then X is referred to as rough with respect to B. The concept of rough sets is usually applied in rough sets based classifiers to extract useful data from decision tables with inconsistent data and to reduce attribute sets [5].

2.2.1 Rule Extraction - Learning from Examples Module, version 2 (LEM2) Algorithm.

Decision rules are generated from the upper and lower approximations of a decision table. In data classification, there are various methods of extracting decision rules from a decision table. A simple method is to transform the decision table into rules using the condition and decision attribute sets. The focus on the rules is that it has to be extracted with minimal attribute sets. Rule induction algorithms using rough set were first implemented in Learning from Examples System - LERS [6]. The LERS uses examples to generate new rules, it then classifies new examples from the rule sets it had earlier generated.

The LEM2 algorithm for rule induction follows the general scheme of rough set based rule induction and extracts the minimum set of decision rules using minimum attributes sets. From the decision table, a single local covering for each concept is computed by relating the attribute-value pair. These local covering are then converted into the rule sets. LEM2 was employed to generate the decision rules in this research.

2.3. Rule Filter

Rules extracted using rough sets based algorithms example the LEM2 algorithm are usually large in numbers and are difficult to apply in decision making. These large rules limits classification capabilities as some rules are redundant. A rule filter is employed to reduce the number of rules. The quality indices of rules in a set computed by a rule quality function, is used as the basis for filtering the rules. The quality index of each rule determines the strength of the rule based on the rule support and coverage measurements.

2.4. Fuzzy Set Theory (FST) and Similarity Function

Fuzzy set theory (FST) or fuzzy logic is an AI technique employed to deal with data that are incomplete, uncertain and vague. It was invented in 1965 by Zadeh and is used in decision making, and in forecasting data with uncertain values [7]. The elements of a fuzzy set have degrees of membership. These membership degrees are numbers between 1 and 0 defined by membership functions of the fuzzy sets. These membership functions specify the degree to which elements of the universe of discourse belong to a set. The membership functions are set of numbers and can be seen as fuzzy numbers [8]. Employing membership functions mechanism, linguistic values can be assigned to the numeric values. The linguistic values described by

membership functions allow for easy classification and generation of compact rule sets. Similarity in fuzzy matches that occur on the decision side of a fuzzy rule is obtained to determine the fuzzy-rule similarity function. This function generates a value between 0 and 1. The similarity function provides the maximum value of the intersection of two fuzzy values and a high value indicates greater similarity.

Some basic FST definitions are described below:

1. Let *X* be a universe of discourse and let x be an element of *X*. A fuzzy set *A* in *X* is defined as [7]:

$$A = \{(x, \mu A(x)) | x \in X\} \tag{5}$$

$\mu A(x): X \rightarrow [0, 1]$ is called the membership function for fuzzy set *A* and indicates elements of x that are members or non-members of *A*, mapping each element to a membership value between 0 and 1. When $\mu A(x) = 1$, x is a member of *A*, and when $\mu A(x) = 0$, x is declared a non-member of *A*.

2. Let *X* be a *universe of discourse* with two fuzzy subsets *A* and *B* having membership functions μ_A and μ_B, respectively. The union of the fuzzy subsets *A* and *B* is denoted as $A \cup B$, and is defined by the equation [7]:

$$\mu A \cup B(x) = \max(\mu A(x), \mu B(x)) \tag{6}$$

The intersection of fuzzy sets A and B, ($A \cap B$) is defined by [7]:

$$\mu A \cap B(x) = \min(\mu A(x), \mu B(x)) \tag{7}$$

3. The *fuzzy subsethood* is a measure of the degree in which a set *A* is a subset of another set *B*. For a *universe of discourse X*, with two fuzzy sets *A* and *B* having membership functions μ_A and μ_B respectively, the *fuzzy subsethood* is defined by [9]:

$$S(A,B) = \frac{M(A \cap B)}{M(A)} = \frac{\sum\limits_{x \in X}^{min(\mu A(x), \mu B(x))}}{\sum\limits_{x \in X} \mu A(x)} \tag{8}$$

Where $S(A,B) \in [0, 1]$

2.5. Discretization Method – Minimize Entropy Principle Approach (MEPA)

Discretization is the process of partitioning continuous variables into a number of sub-ranges and treating each sub-range as a category. MEPA is an entropy-based discretization method used in this research because it uses class information to partition features. In rule-based classification systems, MEPA performs well reducing data dimension and is employed to obtain the membership functions of attributes. Before data can be divided into membership functions, thresholds among the data classes must first be established. An entropy minimization screening method is used to determine the threshold line, after which the segmentation process begins. The first partition process is to divide the data into two classes to obtain the primary threshold which is the threshold value with the lowest entropy value. Applying the threshold value calculations to further subdivide the dataset will result in obtaining the secondary and tertiary threshold values [10]. To find threshold values for a sample ranging between x_1 and x_2, the regions can be represented as $[x1, x]$ and $[x, x2]$, denoted as *p* and *q* respectively, and entropy equation can be written for these regions. The following equations proposed by Christensen [11] are used to determine the entropy value of **x**.

$$S(x) = p(x)Sp(x) + q(x)Sq(x) \tag{9}$$

$$Sp(x) = -[p1(x)\ln p1(x) + p2(x)\ln p2(x)] \tag{10}$$

$$Sq(x) = -[q1(x)\ln q1(x) + q2(x)\ln q2(x)]$$

$$p(x) + q(x) = 1 \tag{11}$$

Where:

$pk(x)$ and $qk(x)$ are probability measures that the class k sample is in the region $[x1, x1 + x]$ and $[x1 + x, x2]$, respectively,

$p(x)$ and $q(x)$ are probability measures that all samples are in the region $[x1, x1 + x]$ and $[x1 + x, x2]$ respectively.

The entropy estimates of a class k sample $pk(x)$, $qk(x)$, and $p(x)$, $q(x)$ are given below:

$$pk(x) = \frac{nk(x) + 1}{n(x) + 1} \tag{12}$$

$$qk(x) = \frac{Nk(x) + 1}{N(x) + 1} \tag{13}$$

$$p(x) = \frac{n(x)}{n} \tag{14}$$

$q(x) = 1 - p(x)$ (15)

Where:

$nk(x)$ = samples of class k in region $[x1, x1+x]$,

$n(x)$ = the total number of samples in region $[x1, x1+x]$,

$Nk(x)$ = samples of class k in region $[x1+x, x2]$,

$N(x)$ = the total number of samples in region $[x1+x, x2]$, and

n = number of all samples in $[x_1, x_2]$.

The primary (PRI) threshold value is the value of **x** in the region with the minimum entropy value. Using the entropy equations and repeating the partition process, secondary (SEC) threshold values and tertiary (TER) threshold values can be determined.

III. Methodology

3.1. Concept of the Proposed Hybrid Model

This research proposes a hybrid model for forecasting abandon rate and to improve the classification quality of a rough sets based classifier. The proposed model consists of the following five AI tools: experiential knowledge (EK), Minimize Entropy Principle Approach (MEPA), fuzzy set theory (FST), rough set theory (RST), and rule filter (RF). A practical dataset (called the AR dataset) collected from Airtel, from 1 December, 2012 – 28 February, 2013 is used to illustrate the proposed model. There are in total, 90 records in the experimental dataset, which are characterized by eight attributes. The methodologies to be adopted in this research involves relating call center in the telecommunication industry to select attributes and then using AI techniques for classifying and forecasting abandon rate. These methodology steps are summarized as follows:

i.. Collection of the attributes of daily inbound calls from Airtel call center.

ii. Preprocessing the dataset to eliminate redundancy.

iii. Partition of data using MEPA to obtain membership functions.

iv. Assigning unique corresponding linguistic values to each attribute of the data.

v. Finding the closeness between the decision and each subgroup of condition attributes.

vi. Transforming the fuzzified attributes into binary numbers.

vii.Extracting decision rules using the RST-LEM2 algorithm.

viii. Filtering the extracted rules to improve the rule quality.

ix. Forecasting and validating the hybrid model.

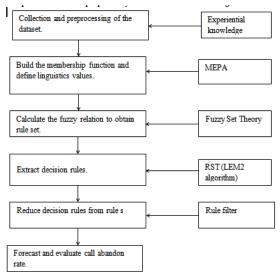The procedures of the proposed hybrid model are shown in Fig. 1.

Figure 1 Block Diagram of the Hybrid Model.

The procedures for the proposed hybrid model are described in details as follows:

Step 1. Collection of the attributes from the call center daily inbound calls by experiential knowledge. The pertinent attributes from the daily inbound calls over a three month period were obtained from Airtel. Table 1 shows the attributes information.

Table 1 Information on Attributes in AR Dataset

No	Attribute	Value Type	Notes
1	Calls Offered (CO)	Continuous	Min: 57804 ~ Max: 214765
2	Calls Answered (CA)	Continuous	Min: 26754 ~ Max: 143670
3	Calls Abandoned (CAb)	Continuous	Min: 2900 ~ Max: 104613
4	Average Talk Time (ATT)	Continuous	Min: 103 ~ Max: 184
5	Average Waiting Time (AWT)	Continuous	Min: 12 ~ Max: 97
6	Number of Agents (Ag)	Continuous	Min: 430 ~ Max: 592
7	Service Level (SL)	Continuous	Min: 36.67 ~ Max: 96
8	Abandon Rate (AR) Class	Continuous	Min: 7.01 ~ Max: 73.65

From Table 1, the data in each attribute are continuous and non-linear. The condition attributes are the first seven attributes, and the abandon rate is the decision attribute, (also known as class). Based on experiential knowledge, the abandon rate (class) is partitioned into three classes:

i. Low (AR ≤ 20%),
ii. Fair (AR is between 20.01% – 49.99%), and
iii. High (AR ≥ 50%).

Step 2. Preprocessing the dataset: In order to select the relevant attributes required for abandon rate and to simplify classification rules, missing values and records with inconsistent and redundant data values are to be removed.

Step 3. Determination of membership function using MEPA: MEPA is applied to partition the continuous attributes and build membership functions. The entropy equations proposed by Christensen is used to compute entropy value of each attribute. Five thresholds are obtained for each of the condition attributes and they are TER1, SEC1, PRI, SEC2, and TER4. Tertiary thresholds TER1 and TER4 are represented as MIN and MAX respectively in this research. Table 2 shows threshold values for the condition attributes in the AR dataset.

Table 2 Threshold Values for Condition Attributes

Attribute	MIN	SEC1	PRI	SEC2	MAX
CO	57804	98781	123069	143451	214765
CA	26754	54668	76990	80475	143670
CAb	2900	18639	20891	65619	104613
ATT (sec)	103	130	137	138	184
AWT(sec)	12	43	49	53	97
Ag	430	480	490	522	592
SL (%)	37	62	66	73	96

Step 4. Fuzzification of the condition attributes into linguistic variables: The membership functions obtained using MEPA were used to grade data into membership degrees. The linguistic value for each of these data within an interval was determined by the maximal membership degree in that interval. The triangular membership function which is a collection of straight lines and three points forming a triangle was used. The discretized data are converted into linguistic values through this process. Three linguistic values were obtained for the condition attributes using the threshold range. They are: L1 – small, L2 - medium, and L3 – large.

Each data in an attribute will belong to a linguistic value (L1, L2 or L3). Fig. 2 shows the graphical plot of the membership functions of the 'CO' attribute using threshold values obtained by the MEPA. The value '1' is assigned as the membership degree when the attribute's value is lower than the SEC1 threshold and greater than the SEC2 threshold values.
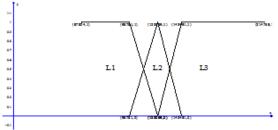

Figure 2 Membership Functions of the 'CO' Attribute.

Step 5. Determining subsethood values between attributes by Fuzzy Set Theory (FST): The fuzzy subsethood value is computed to find the degree in which the decision attribute (denoted as A) is a member of each subgroup of the condition attributes (denoted as B). The decision attribute being classified into three classes; Low (L), Fair (F) and High (H), gave rise to three subgroups; subgroup1, subgroup2 and subgroup3 respectively. The subsethood degree of fuzzy sets can be defined as the ratio:

$$S(A,B) = \frac{\sum_{x \in X} min(\mu A(x), \mu B(x))}{\sum_{x \in X} \mu A(x)}$$

(16)

The similarity threshold of linguistic values for all condition attributes is computed using the equation:

$$S_T = \max \left\{ S(C_s, L_i) \right\}$$

(17)

Where:

S_T is the similarity threshold, and

$S(C_{s, Li})$ is the subsethood of same linguistic value in a subgroup.

S_T is determined by selecting the maximal value of subsethood comprising of same linguistic value for the three classes. Table 3 shows the similarity thresholds for the condition attributes in the AR dataset.

Table 3 Similarity Thresholds of the Condition Attributes

No	Attribute	L1	L2	L3
1	CO	0.700	0.325	0.413
2	CA	0.672	0.379	0.784
3	CAb	0.797	0.530	0.747
4	ATT	0.931	0.173	0.778
5	AWT	0.853	0.106	0.833
6	Ag	0.422	0.102	0.937
7	SL	0.868	0.197	0.869

Step 6. Transformation of the fuzzified attributes: The membership degree obtained for each linguistic value of a condition attribute was compared with the similarity threshold obtained for that same linguistic value of same attribute. Based on this comparison on the fuzzified dataset, they were transformed into binary, i.e. 1's and 0's. The fuzzified dataset are transformed using the comparison process:

$$I'i = \begin{cases} 1, & I_i \geq S_T \\ 0, & Otherwise \end{cases}$$

(18)

Where:

I_i is the linguistic value's membership degree of condition attribute, and

I'_i is the new transformed value after comparison.

Table 4 The Partial AR Dataset after Transformation

No	CO			CA			...	SL		
	L1	L2	L3	L1	L2	L3	...	L1	L2	L3
1	0	1	0	0	1	0	...	0	1	0
2	0	1	0	1	0	0	...	1	0	0
3	1	0	0	0	0	1	...	0	0	1
.	.	.	.	.	.	.	...	.	.	.
.	.	.	.	.	.	.	...	.	.	.
.	.	.	.	.	.	.		.	.	.
89	0	0	1	0	0	1	...	0	0	1
90	0	1	0	0	1	0	...	1	0	0

Step 7. Extraction of decision rules by the rough set based LEM2 algorithm: Records of the transformed dataset are split into two, 67% for training the rules and 33% used for testing the rules. The LEM2 algorithm was used on the training set to extract the decision rules. The decision rules generated are grouped into three classes, which are Low, Fair, and High. Real examples coinciding with the generated decision rules are termed 'Supports'.

Step 8. Filter the extracted rules to improve rule quality using the rule filter mechanism: The quality index of each rule set is determined as the basis for filtering the rules. The rule support and coverage values were also

determined, based on these indices, rules that are below the target threshold value were removed. The rules were filtered in other to eliminate redundant rules and improve classification accuracy. A total of 28 decision rules were generated after filtering.

Step 9. Classify, forecast call abandon rate and validate using obtained data from Airtel. Records of the condition attributes in the testing set now represented in linguistic terms are classified as belonging to a decision class using the rules extracted from the training set. A forecast of the linguistic attributes for future abandon rates was also generated. The center of gravity defuzzification technique was employed to defuzzify the linguistic data and generate numeric forecasts.

IV. Result And Discussion

The decision rules generated were used to classify and forecast the AR dataset obtained from Airtel, and the performance of the hybrid model was evaluated and described in this section. The proposed hybrid model's performance was verified for its classification accuracy, and the hybrid model was also compared with the C4.5 and the traditional rough sets models. The classification accuracy rates, number of rules generated, and standard deviation were used as criteria for evaluating the proposed hybrid model.

4.1. Comparison

For verification purpose in cases of classification systems the quality of the set of rules is evaluated using the classification accuracy rate. The records of the training and testing sets were classified as 'Low', 'Fair' or 'High', using the 28 generated decision rules. The percentage of successful classification was used to measure classification accuracy rate of the hybrid model on attributes in the AR dataset. Records of the training and testing set were grouped based on the generated rules. The percentage of successful classification (PSC) is expressed in percentage and is defined by [12]:

$$PSC = \frac{TP}{TP + TN} \times 100 \qquad (19)$$

Where:
TP = Group correctly classified, and
TN = Group wrongly classified

4.1.1. Classification Accuracy of Training Set

The first comparison result was obtained from the 67% dataset used for training. All records in the training set were classified, using the 28 decision rules generated. The rules were used to perform predictions and classified the dataset as belonging to a certain decision class. The result for the internal comparison of the training set is shown in Table 5.

Table 5. Classification Accuracy of Training Set in the AR Dataset

Class Members	Group Correctly Classified	Group Wrongly Classified
Low (15)	15 (100%)	0
Fair (31)	31 (100%)	0
High (14)	14 (100%)	0

The PSC for the training set is 100%

From Table 5, all 15 records classified as 'Low' were correctly predicted using the generated rules by the hybrid model. There are 31 records classified as 'Fair' and 14 records classified as 'High' and they were all correctly classified by the rules. The internal comparison results thus gave a PSC of 100%.

4.1.2. Classification Accuracy of Testing Set

For testing classification rate of the hybrid model, the 33% testing set from the AR dataset was used. The rules were used to classify the set as belonging to a certain decision class. The result of the internal comparison for the testing set is shown in Table 6.

Table 6 Classification Accuracy of Testing Set in the AR Dataset

Class Members	Group Correctly Classified	Group Wrongly Classified
Low (10)	10 (100%)	0
Fair (16)	14 (87.50%)	2 (12.50%)
High (4)	3 (75.00%)	1 (25.00%)

The PSC for testing set is 90.00%

From Table 6, all ten records are members of class 'Low' There are 16 records in the 'Fair' class, and four records in the 'High' class. The testing set had 27 correctly classified records and three wrongly classified records giving a PSC of 90%.

4.1.3. External Comparison
The hybrid model was compared with two AI models; the C4.5 and the traditional rough sets (TRS) models using the AR dataset to compare its performance. This comparison was done on the testing set using decision rules generated from the various models and the seven condition attributes partitioned into three linguistics variables (L1, L2 and L3). Table 7 shows the performance comparison for the proposed hybrid, the C4.5 and the TRS models.

Table 7 External Comparison Results

Model	Training Accuracy(%)	Testing Accuracy(%)	Rules
C4.5	100 (0.0)	86.67 (16.73)	32
TRS	100 (0.0)	76.67 (15.69)	74
Proposed Hybrid	100 (0.0)	90.00 (15.08)	28

As seen in Table 7, the proposed hybrid model has the highest classification accuracy rate and the lowest standard deviation. The proposed model was found to be more accurate when compared to the other methods by an accuracy of 90% using the minimal set of reduced rules, and it generated 28 decision rules, which is the least number of rules. Based on these evaluations, it is obvious that the hybrid model outperformed the other AI techniques. To further evaluate efficiency of a hybrid model over a stand-alone model, the improved accuracy rate and improved rule rate was determined for the proposed model and the TRS method. The result is shown in Table 8.

Table 8 The Improved Accuracy and Rule Rates for the Proposed Hybrid and TRS Models

Model	Testing Accuracy (%)	Improved accuracy rate (%)	Rules	Improved rule rate (%)
TRS	76.67	-	74	-
Proposed Hybrid	90.00	17.39	28	62.16

From Table 8, the proposed hybrid model had higher accuracy and rule rates than the traditional rough sets model. It is a better method to forecast abandon rate and the result demonstrates that the traditional rough sets method in forecasting still needs to be improved.

4.2. Result Analysis
This research proposed a hybrid model which integrated five AI tools to enhance forecasting accuracy. The proposed model outperformed the other testing models using some criteria as seen in Table 7, and from validation of these results, the following analysis were made:
- Accuracy rate: The proposed hybrid model had the highest accuracy rate of 90% while the C4.5 and the traditional rough sets models have accuracy rates of 86.67% and 76.67% respectively.
- Standard deviation on accuracy rate: The proposed hybrid model had the lowest standard deviation value of 15.08, indicating that the data points tend to be closer to the expected value. While the C4.5 and the traditional rough set models have standard deviation values of 16.73 and 15.69 respectively.
- Number of generated rules: The proposed hybrid model generated a number of 28 filtered rules, which is lesser than the other models.
- Number of attribute sets used in rule extraction: During decision rules generation, the proposed model used the least number of attribute sets.
- An effective way for fuzzy sets and rough sets combination methods in forecasting: Results from the improved accuracy and improved rule rates indicates that the hybrid models surpassed the stand-alone models.

V.   Conclusion
The research has proposed a hybrid model constituting of five various AI components: experiential knowledge, MEPA, fuzzy set theory, rough set theory and rule filter for classifying and solving the problems of forecasting call abandon rate faced by operator. The decision rules extracted from the model were used to

classify attributes of the experimental dataset. The hybrid model was compared with the C4.5 and the traditional rough sets models, using the classification accuracy rates, the number of rules generated, and standard deviation as basis for evaluation. The experimental results showed that the proposed hybrid model had the highest accuracy of 90%, the lowest standard deviation of 15.08 and generated the lowest number of rules (28) using minimum set of attributes. Thus, it outperformed the other AI models. Rule sets extracted from the hybrid model also supported the 'IF - THEN' rule form to predict abandon rate, which is an important evaluation indicator for telecommunication operators.

Acknowledgments

References

[1].   H. Shu-guang, L. Li. & Q. Er-shi, Study on the Continuous Quality Improvement of Telecommunication Call Centers Based on Data Mining. Service Systems and Service Management, 2007. International Conference, 1-5.

[2].   H. Shen, Statistical Analysis of Call-Center Operational Data: Forecasting call arrivals, and analysing customer patience and agent service, 2010. Wiley Encyclopedia of Operations Research and Management Science.

[3].   A. Ravi, H. Kurniawan, P.N.K. Thai and P.K. Ravi. Soft computing system for bank performance prediction. 2008. Applied Sot Computing, vol. 8, pp. 305-31.

[4].   M. Inuiguchi and M. Tsurumi, Measures based on upper approximations of rough sets for analysis of attribute importance and interaction, 2006. International Journal of Innovative Computing, Information and Control, vol.2, no.1, pp.1-12.

[5].   Z. Pawlak, Rough Sets-Theoretical Aspects of Reasoning About Data, 1991. Dordrecht: Kluwer Academic Publishers. ISBN 0792314727.

[6].   J.W. Grzymala-Busse , A new version of the rule induction system LERS. 1997. Fundamenta Informaticae, vol.31, no.1, pp.27-39.

[7].   L. Zadeh. Fuzzy Sets. 1965. Information and Control, pp.338-353.

[8].   L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning, 1975. Information Science, vol. 8, pp.119-249.

[9].   Y. Yuan and M.J. Shaw, Induction of fuzzy decision trees, Fuzzy Sets and Systems, 1995. vol. 69, pp.125-139.

[10]. T.J. Ros. Fuzzy logic with engineering applications. 2000. International edition, McGraw-Hill, USA.

[11]. R. Christensen. Entropy Minimax Sourcebook, 1980. Philosophical Origins. Entropy Ltd., Lincoln, MA. Vol II, ISBN 0-938876-04-X

[12]. A.O. Adetunmbi, A. Oguntimilehin and S.O. Falaki. Web-based Medical Assistant System for Malaria Diagnosis and Therapy, 2012. GESJ: Computer Science and Telecommunications, Vol.33, No. 1, ISSN 1512-1232.